経験類似度に基づくボラティリティ予測

川崎　能典　モデリング研究系　教授

1. 高頻度データとRV, RQ

秒単位で記録された金融取引データを1分の等間隔データに直したものを$r_{t,i}$と記す．$r_{t,i}$は第t日におけるi番目の収益率(価格の対数差分)である．n_tを第t日の標本数とするとき$\mathrm{RV}_t = \sum_{i=1}^{n_t} r_{t,i}^2$を第$t$日のrealized volatility (RV)と呼ぶ．また，収益率の4次モーメントに対応する量として，realized quarticity (RQ)を$\mathrm{RQ}_t = (n_t/3)\sum_{i=1}^{n_t} r_{t,i}^4$で定義しておく．$\mathrm{RV}_t$の推定誤差に対応する量である．

2. 類似度に基づくボラティリティ予測

日次のボラティリティRV_tを以降v_tと記す．v_tを観測値と便宜的に見なし，v_tに対する予測モデルを立てる．日次ボラティリティの週次平均と月次平均を$v_t^{(w)} = 5^{-1}\sum_{i=1}^{5} v_{t-i+1}$ならびに$v_t^{(m)} = 22^{-1}\sum_{i=1}^{22} v_{t-i+1}$とする．ここで，1期前の日次，週次，月次ボラティリティと当期のボラティリティとの類似度から決まる重みを

$$\begin{aligned}\theta[v_{t-1}, v_{t-2}] &= \exp\left(-\omega_1(v_{t-1}-v_{t-2})^2\right)\\ \theta[v_{t-1}, v_{t-2}^{(w)}] &= \exp\left(-\omega_2(v_{t-1}-v_{t-2}^{(w)})^2\right)\\ \theta[v_{t-1}, v_{t-2}^{(m)}] &= \exp\left(-\omega_3(v_{t-1}-v_{t-2}^{(m)})^2\right)\end{aligned}$$

とし，

$$v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(w)}]v_{t-1}^{(w)} + \theta[v_{t-1}, v_{t-2}^{(m)}]v_{t-1}^{(m)}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(w)}] + \theta[v_{t-1}, v_{t-2}^{(m)}]} + \epsilon_t$$

というモデルを考える($\epsilon_t \overset{iid}{\sim} (0, \sigma^2)$とする)．これをES1モデルと呼ぶ．これを単純化し，$\theta[v_{t-1}, v_{t-2}] = \theta[v_{t-1}, v_{t-2}^{(w)}] = \theta[v_{t-1}, v_{t-2}^{(m)}] = 1/3$と制約したモデルをES0モデルと呼ぶ．更に，Corsi (2009, *J. Financial Econometrics*)のHeterogeneous AutoRegressionモデル(HARモデル)

$$v_t = \alpha_0 + \omega_1 v_{t-1}^{(d)} + \omega_2 v_{t-1}^{(w)} + \omega_3 v_{t-1}^{(m)} + \varepsilon_t, \quad \epsilon_t \overset{iid}{\sim} (0, \sigma^2)$$

も比較の俎上に乗せる．

3. RQによるモデル拡張

v_tを自己回帰モデルで予測する際に，係数をRQ_tに依存させることで時変係数モデルとすることが近年注目されている．例えば

$$v_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} v_{t-1} + \epsilon_t$$

とする定式化は，AutoRegressive Quarticity (ARQ) モデルと呼ばれる．同様にHARモデルに対しても

$$v_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} v_{t-1} + \beta_2 v_{t-1}^{(w)} + \beta_3 v_{t-1}^{(m)} + \epsilon_t$$

と拡張できる．このモデルは，Heterogeneous AutoRegression Quarticity (HARQ) モデルと呼ばれる．

ところで，先に述べたESモデルの構成方法は，ボラティリティ予測子の組合せに対して柔軟に適用することが可能である．そこで，過去の日次ボラティリティv_{t-1}，HAR予測子$v_{t-1}^{(\mathrm{har})}$，および上述のHARQ予測値$v_{t-1}^{(\mathrm{harq})}$を組み合わせた新しいモデルを考えることも可能となる．

$$v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(\mathrm{har})}]v_{t-1}^{(\mathrm{har})} + \theta[v_{t-1}, v_{t-2}^{(\mathrm{harq})}]v_{t-1}^{(\mathrm{harq})}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(\mathrm{har})}] + \theta[v_{t-1}, v_{t-2}^{(\mathrm{harq})}]} + \epsilon_t$$

これをESQと呼ぶ．類似度関数の構成はES1と同様である．

4. 予測力比較

(データ) 用いたデータは1999年1月4日から2013年12月30日までの15年分の株価指数と，流動性のある個別銘柄(大型株)である．予測比較では，インサンプル，アウトオブサンプルを含む計225の推定予測期間を分析対象とする．詳細な設定は，森本・川崎 (2017)を参照．

(誤差関数) 実測値v_tと予測値$\hat{v}_t$の距離は，Patton (2011, *J. Econometrics*)の誤差関数$\mathcal{L}$で測る．紙幅の都合で$\mathcal{L}$の定義は森本・川崎 (2017)を参照．

(モデル信頼集合) t時点におけるi番目のモデルの予測値$\hat{v}_{i,t}$の当てはまりの尺度$\mathcal{L}(v_t, \hat{v}_{i,t})$と，$j$番目のモデルでの$\mathcal{L}(v_t, \hat{v}_{j,t})$との差に統計的に有意な差があるかどうかを，ブートストラップに基づく対比較(検定)で行う．同等性を表す帰無仮説が棄却されると，劣ったモデルは候補集合から外されていく．最終的に複数のモデルが残る場合もあるが，それらは同程度に最良予測モデルを与えると見なす．このようにモデル信頼集合を形成する方法は，Hansen et al. (2011, *Econometrica*)で提案されており，本研究ではそれに従う．

(予測結果と解釈) 下の棒グラフに，各モデルが最良予測モデルとして残った頻度がまとめられている．株価指数(Index)に対しては，内挿ではESQとHARQが高い予測能力を示しているが，外挿となるとESQが傑出してよい．個別銘柄(Individual)に対しては，内挿ではHARQが最も良く，次いでESQ，ES1である．個別銘柄の外挿では，ノンパラメトリック予測であるES0が最も予測能力が高いという結果を得た．株価指数に比べ，個別銘柄ではRQの効きは弱い．これは，個別銘柄の方が一時的な変動が大きく，過去の履歴からRVの変動性を把握していても，系統的に予測の役に立つ割合は株価指数より少ないからであると考えられる．

参考文献　森本孝之，川崎能典 (2017). 経験類似度に基づくボラティリティ予測,『統計数理』第65巻1号, 155–180.